Statistical Analysis Plan

<u>Administrative information</u>
1. **Title and trial registration**
   Statistical analysis plan (SAP) for the quantitative outcome measures of the TeamUp intervention cluster randomized control trial (cRCT) in Burundi
   Trial registration: International Society for the Registration of Clinical Trials, ISRCTN17499603. Registered on 19 October 2023: https://doi.org/10.1186/ISRCTN17499603

2. **SAP version**
   This statistical analysis plan was written in adherence to the guidelines for SAPs in clinical trials (Gamble et al., 2017).
   First version, October 2024
   Last version, December 2024
   Signed version, January 2025

3. **Protocol version**
   Version 1.0 submitted to the National Institute for Public Health (Institute National Santé Publique, INSP) in Burundi on 8 August 2023. Ethical approval, called "Visa Statistique" was initially obtained on 18 August 2023 up to 17 August 2024. The protocol and approval from INSP was also submitted to the National Institute of Statistics of Burundi (Institut National de la Statistique du Burundi, INSBU) on 20 August 2023 for final clearance.

   Version 1.1 was submitted – after making amendments to the protocol and to extend the needed time frame to cover the period of the planned study – on 6 August 2024 to INSP and on 20 August to INSBU. The "reviewed" ethical approval, "Visa Statistique" was obtained to cover 14 August 2024 for up to 13 August 2025.

4. **SAP revisions**
   none

5. **Roles and responsibilities**
   Alexandra Bleile, MSc and PhD-candidate, Research and Development, War Child Alliance, Amsterdam, The Netherlands: Lead researcher.
   Gabriela Koppenol-Gonzalez, PhD, Research and Development, War Child Alliance, Amsterdam, The Netherlands: Lead statistician.
   Jean Bosco Mutima, MA, Research and Development, War Child Burundi, Bujumbura, Burundi: Research Coordinator.
   Ioannis Bakolis, PhD, Institute of Psychiatry, Psychology, and Neuroscience at King's College London, London, UK: Senior statistician consultant.
   Renovate Irambona, Professor, Faculty of Psychology and Educational Sciences, University of Burundi, Bujumbura, Burundi: Co-Investigator.
   Mark Jordans, Professor, Research and Development, War Child Alliance, Amsterdam, and Amsterdam Institute of Social Science Research, University of Amsterdam, Amsterdam, Netherlands: Principal Investigator and Director of Research and Development at War Child Alliance.

6. **Signatures**

Name	Date	Signature
Alexandra Bleile	20/01/2025	*[signature]*

Gabriela Koppenol-Gonzalez	20/01-2025	
Jean Bosco Mutima	16/01/2025	
Ioannis Bakolis	15/01/2025	
Renovate Irambona	16/01/2025	
Mark Jordans	19/01/2025	

Introduction

7. Background and rationale

TeamUp is a psychosocial movement-based, mental health promotion intervention, with a non-verbal and embodied-learning approach, consisting of a wide variety of movement-based group games, sports, creative movement, routines, body awareness and breathing exercises. Non-specialist, trained adult facilitators follow a guiding structure, creating a non- judgmental and safe space, where children are invited and encouraged to try new activities and ways of moving their bodies (TeamUp, 2020a). Underpinned by; (i) the Inter-agency Standing Committee Mental Health and Psychosocial Support Guidelines for Emergencies (Inter Agency Standing Committee, 2007), (ii) trauma-informed care principles (Hobfoll et al., 2007), (iii) body-mind interconnection conceptualizations (Bareka et al., 2019; Dieterich-Hartwell, 2017; Dieterich-Hartwell & Koch, 2017), and (iv) activities that promote self-calming, self-regulation and social interaction (TeamUp, 2020a; Pierce, 2014; Tortora, 2005). TeamUp sessions follow a guiding structure – (i) opening routine and check-in, (ii) a body warm-up, (iii) main activities (iv) a cooling-down (v) a check-out and closing routine, and facilitators use four basic facilitation techniques—flow, build-up, group organization and demonstration.

Sessions are based on eight psychosocial themes that are addressed through specific behaviour and observable skills (i.e. fear, anger, respect, conflict, bullying, friendship, stress & tension, and assertiveness – see Annex A). The weekly 45-60 minute sessions are offered by a team of 3–5 trained facilitators for children of a specific age group, e.g. 9-15 year-olds (if in schools, commonly with an age range of one specific year/grade).

From 2019-2021 we completed a quasi-experimental study in Uganda demonstrating TeamUp to have promising outcomes: significantly improving children's emotional and psychosocial wellbeing (M_{diff} = −1.49, SE = 0.6, p = .01), satisfaction with and attitude toward school (−0.57, SE = 0.2, p = .004); as well as (secondary outcomes), showing significantly improving children's traumatic stress (2.64, SE = 0.8, p < .001), health-related quality of life (−1.56, SE = 0.4, p = .001), physical health (−0.78, SE = 0.3, p = .014) and the TeamUp mechanisms of action scale (−3.34, SE = 0.9, p < .001), specifically the subscales social connectedness (−0.74, SE = 0.3, p = .007) and sense of agency (−0.91, SE = 0.3, p = .005), compared to the control group (Bleile et al., 2024). Most results were sustained at 5 months post- intervention follow-up, and mediation analyses demonstrated that children's increased sense of connectedness at endline mediates the effect of TeamUp on improving well-being at 5 months follow-up (indirect effect = 0.30, SE = 0.13, p = .001), explaining 15% of variance (Jordans et al., 2024).

Although the TeamUp intervention was not designed with a clear beginning and end point in mind, but rather as a drop-in/drop-out option for children in asylum seeker/refugee reception centres in the Netherlands (TeamUp, 2020a; Bleile et al., 2021), establishing a specific number and optimal duration of sessions is crucial for evaluating effectiveness, determining cost-effectiveness and informing ongoing implementation practice in diverse countries as well as global scaling efforts through partners.

A fully-powered three-armed cluster Randomised Controlled Trial (cRCT) enables us to test the effectiveness of the TeamUp intervention on children's psychosocial wellbeing, as well as examine the optimal dosage, mechanisms of change, long-term effect (6 months follow up) and cost-effectiveness.

8. Objectives
The primary aim of this 3-armed cluster randomised control trial (cRCT) is to; (i) evaluate the effectiveness of the TeamUp intervention for children aged 9-15 to improve the psychosocial wellbeing of children compared to the control group. Given the 3-armed design, the co-primary aim is to; (ii) to evaluate the impact of differential dosage of the TeamUp intervention (control, 12 TU sessions and 24 TU sessions) on children's psychosocial wellbeing.

Secondary objectives include; (iii) the evaluation of the effects of TeamUp on secondary outcomes, including traumatic stress, attitudes towards school, friendships, quality of life/subjective wellbeing, physical health/wellbeing, self-defined problems, hope, emotional regulation, depression symptoms, and TeamUp Mechanisms of Action scale (sense of safety, social connectedness, sense of agency, emotional regulation and interoception); (iv) to evaluate sustained effects of TeamUp, on primary and secondary outcomes after 6 months follow up to assess whether after 6 months there is a sustained difference between the three arms and particularly between children having attended 12 TeamUp sessions and those having attended 24 TeamUp sessions. Furthermore, we aim (v) to evaluate the cost-effectiveness of the TeamUp intervention, i.e. the cost of the intervention per improved wellbeing outcome (at endline). The cost-effectiveness analyses are outside of the scope of this SAP and are stipulated separately.

Furthermore, we will evaluate the caregiver-reported outcomes (wellbeing of children), assessed at three time points. We will also examine teacher-reported outcomes (classroom dynamics in general and behaviour of individual (index) children), as well as facilitator wellbeing, resilience, satisfaction and burden, assessed pre- and post- intervention. This will enable us to evaluate spillover effects of the intervention – beyond the wellbeing of children – on school-related, teacher-reported outcomes as well as on the service provider (facilitator) outcomes.

Primary and co-primary hypotheses are:
1) Children in the TeamUp (12 sessions) have significantly better psychosocial outcomes compared to the control group at endline
2) Children in the TeamUp (24 sessions) have significantly better psychosocial outcomes compared to the control group at endline
3) Children in the TeamUp 'higher dosage group' (24 sessions) have significantly better psychosocial outcomes compared to the children in the TeamUp 'lower dosage group' (12 sessions) at endline

Secondary hypotheses:
1) Children who participated in TeamUp have significantly better psychosocial wellbeing at 6-month follow-up (i.e. 6 months after the end of session implementation), compared to the control group

2) Mechanisms of Action, hope and emotional regulation will mediate the change in psychosocial wellbeing as a result of receiving TeamUp.
3) While we hypothesise above that 24 TeamUp sessions are more effective than 12 TeamUp sessions, we hypothesise that 12 TeamUp sessions will be more cost-effective compared to 24 TeamUp sessions (not included in this analysis plan).
4) The implementation of TeamUp sessions, thus the engagement of facilitators in the psychosocial movement-based activities while providing them to the children, has a spill-over effect and improves the psychosocial wellbeing of facilitators (service providers) themselves, based on facilitator-reported outcomes
5) Children's behaviour in the classroom and the teacher-student relationship improves as a result of TeamUp over time and compared to the control group classrooms, based on teacher-reported outcomes
6) Children's psychosocial wellbeing based on caregiver-reported outcomes improves as a result of TeamUp over time and compared to the control group.

Study Methods
9. Trial design
The 3-armed cluster randomised control trial, comparing children participating in 12 TeamUp sessions, with children participating in 24 TeamUp sessions and children joining activities as usual (control group).

10. Randomization
We randomly selected 36 schools out of the 42 eligible schools (following the mapping and selection of eligible schools). We randomized at the level of schools rather than at the level of individual children, due to the nature of the intervention (group based), practical and ethical reasons. In this study, statistical methods will be applied to account for clustering of children at school level.

The statistician randomly created three groups, thus lists of schools A, B, C. During a meeting with local school authorities, the groups A, B, and C were matched to one of the three study conditions (12 TeamUp sessions, 24 TeamUp sessions and control group schools) with school authorities picking a paper per trial arm (labelled with the three conditions) out of a box. The allocation of A, B, C was written on a flip chart for everyone to see. This ensured buy-in, transparency and support of the study by local school authorities and school directors. Only after the meeting, school directors were informed individually whether their school would be part of group A, B, and C, in order to reduce unmasking and reduce disappointment during the meeting.

As facilitators needed to be resident within a feasible distance to (three) schools in which they would implement TeamUp, and school directors needed to be aware of the recruitment process within their communities for support and buy-in, randomisation was done *before* facilitator recruitment and training, and before baseline. This enabled start of TeamUp session implementation right after completion of baseline data collection – which otherwise would have been delayed.

Given the nature of the intervention (sessions taking place in the school playgrounds) it will not be possible to mask (i) participants, (ii) TeamUp facilitators, (iii) TeamUp implementation staff (the TeamUp coordinator, two TeamUp mentors, and programme support staff) in Burundi. The following people will remain masked; (i) Principal Investigators and co-Investigators (except the Research Coordinator and the Scientific Coordinator who are responsible for the daily management of the trial); (ii) Lead Statistician and Senior statistician consultant; (iii) Research

Assistants – until the main analyses are completed. Masking of Research Assistants will be ensured by ensuring that Research Assistants and TeamUp facilitators are trained separately and will not meet during the course of the research study. They will have different meeting places. Research Assistants will be masked throughout the study and possible unmasking will be assessed at midline, endline and follow-up.

11. Sample size

The study took place in 36 schools, with 12 schools in each condition – all within Nyanza Lac and Mabanda communes within Makamba Province, southern Burundi. At each school, 60 children on average (ranging from 40-100 children) of grade 4, participated in the study, as well as their caregivers and the class teacher. We also collected quantitative data from 48 TeamUp facilitators (24 of them providing 12 TeamUp sessions, and 24 of them providing 24 TeamUp sessions) pre and post session implementation.

We have powered the study as a superiority study. A target of 2160 children and their caregivers were recruited, with 720 children in each trial arm. We calculated the sample size based on at 1.7% level of significance and 90% power. We accounted for 20% attrition. We assumed an effect size of Cohen's d=0.2, with the SCWBS at endline for either 12 sessions of TeamUp (vs control) OR 24 sessions of TeamUp (vs control) OR 24 months of TeamUp (vs 12 sessions of TeamUp) accounting for clustering with an intraclass correlation coefficient (ICC) of 0.03 (in the quasi-experimental study this was 0.02), between-arm variance of 4 and between-cluster variance of 1.5.

12. Framework

This main aim of the study has a <u>superiority</u> hypothesis testing framework. We aim to show the superiority of 24 TeamUp sessions compared to 12 TeamUp sessions, compared to control condition on our primary outcome (psychosocial wellbeing) and secondary outcomes at endline.

If no superiority is shown, a <u>non-inferiority</u> test will be employed, assessing if 12 TeamUp sessions are not significantly worse than 24 TeamUp sessions, on the primary outcome as well as secondary outcomes at endline.

We will evaluate whether the Mechanisms of Action scale (including its subscales), hope and emotional regulation, mediate the change in psychosocial wellbeing.

In addition, we will combine the data of 12 sessions and 24 sessions, to compare TeamUp (any amount) to control, hypothesizing the combined TeamUp arms to be superior to the control arm. Further, we will evaluate if attendance is a <u>predictor</u> of outcomes.

With the midline data points (collected for 24 TeamUp session arm and control, at the same time as endline of the 12 TeamUp session arm), we aim to assess the <u>equivalence</u> of the primary outcome (psychosocial wellbeing, SCWBS) – comparing endline of TeamUp 12 sessions with midline of TeamUp 24 sessions. We will also test for any difference between midline of 24 TeamUp sessions with control.

Details are stipulated in section 29.

13. Statistical interim analyses and stopping guidance

We collected pilot data to assess the psychometrics of the child-reported outcome measures in two schools (n=169 children) within the same area of the study, not part of the 36 selected schools.

No interim analyses were conducted for the cRCT and there are no stopping guidelines for the trail.

14. Timing of final analysis
All outcomes will be analyzed collectively after the final data collection, i.e., after 6 month follow-up, and locking of the dataset, see timeline below. The statistician will therefore remain masked to the data until all data are collected, cleaned, and the SAP signed by all co-authors.

15. Timing of outcome assessment
All the outcomes were measured at three time points (see Figure 1 below):
o baseline (October/November 2023)
o endline (immediately post-intervention, thus February/March 2024 for the 12 session children; and May/June 2024 for the 24 session children)
o and 6-months follow-up (counted from endline, thus in September 2024 and November 2024).

At midline for the 24 session and control group children (same timing as endline of the 12 session children) we only assessed the primary outcome (SCWBS) and the PSYCHLOPS (impact of self-defined problems).

Figure 1: timeline
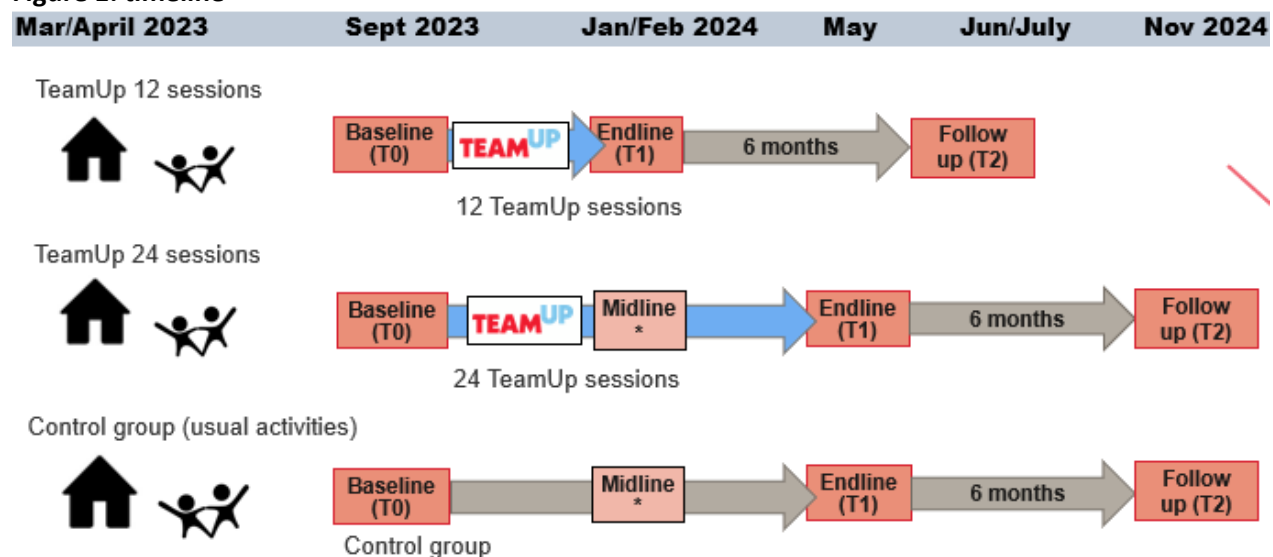
Statistical Principles

16. Level of statistical significance
A significance level of α = 0.05 will be applied.

17. Adjustments for multiplicity
For the primary analyses, we adjust for multiple testing, using the Bonferroni Correction, thus apply a significance level of α = 0.017.

18. Confidence intervals
We will apply 95% CIs for all analyses. For the multilevel analyses 95% CIs will be given for the mean differences between intervention groups and for the mediation analysis 95% CIs will be given for the estimated total, direct, and indirect effects.

19. Fidelity and protocol deviations

Fidelity to the intervention was measured with a 27-item fidelity checklist at individual facilitator (12 items) and team (15 items) level, and in addition 6 items assessing TeamUp-specific competencies. This was administered during mentoring/supervision visits by the mentors/supervisors, about 30-40% of all implemented sessions.

The initial protocol included a block randomization to be done after baseline. Due to community and school authority buy-in needed prior to facilitator recruitment as well as feasibility factors, randomization was done before baseline. Implementation only started after baseline data collection. School holidays were 2 weeks at Christmas and 2 weeks at Easter, in which TeamUp session implementation was not possible. The holidays were the same for all schools in all three arms.

20. Analysis populations

Burundi has been receiving very little attention from the humanitarian community (Ventevogel, 2015; Ventevogel et al., 2011; UNHCR, 2019). A landlocked country in East Africa, Burundi, one of the most densely populated countries in the world and among the poorest worldwide, with 80% of the population employed in agriculture (The World Bank, 2023). Moreover, it has been strongly affected by decades of civil war, conflict and ethnic tensions – following colonial oppression.

We will be studying children and communities affected by conflict and ethnic tensions – in Makamba Province – who are repatriated and returnees (previous refugees in Tanzania or internally displaced within Burundi) and from the local community, with a need for strengthened integration, social cohesion and improved wellbeing. The area is rural, hilly with limited infrastructure.

Given the target age range for TeamUp and age range allowing children to have the cognitive abilities to reply to the questionnaires, we involved children in grade 4, aged 9-15 years within public, primary schools. Studying TeamUp in schools enables individual tracking, as well as potential effect on school attendance, children's concentration and classroom dynamics/behavior in school.

Primary analyses will use an intention-to-treat (ITT) approach, dealing with missing values at any timepoint with multiple imputations (5 datasets). Per Protocol (PP) analysis will also be conducted and is defined as "analysis that considers only the outcomes of participants who strictly adhere to the treatment assigned by randomization and to other protocol requirements (Molero-Calafell et al., 2024)". PP analysis will be conducted in which the primary and secondary outcomes will be compared between groups, including only the children with data at all timepoints (baseline, midline, endline and follow-up) and those that received equal to or 75% of the intended TeamUp sessions. The PP analysis will compare intervention effectiveness as done for the ITT population and will be completed after all other main analyses, as it will unmask the trial statistician.

Trial population

21. Screening data
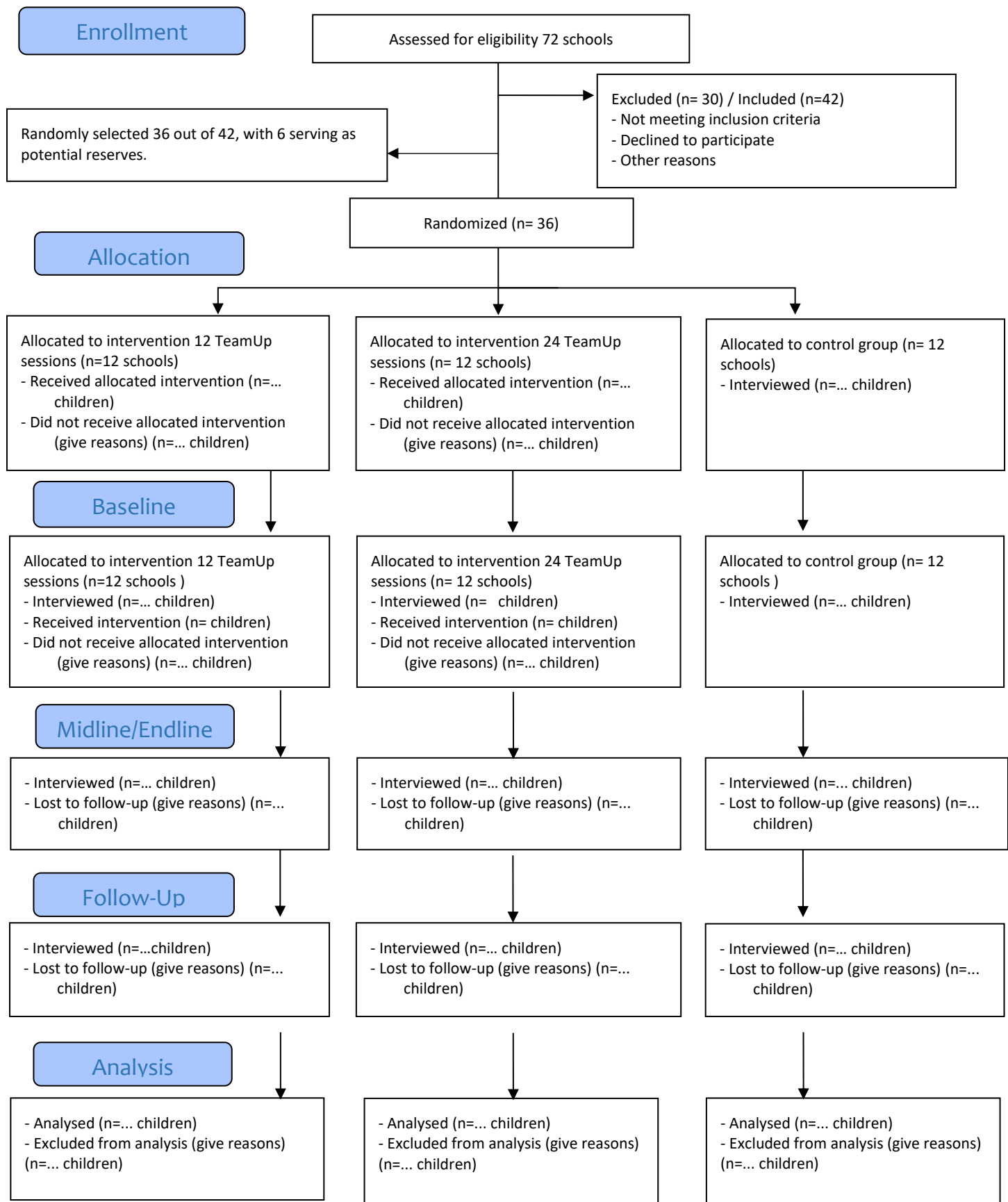There were no diagnostic nor clinical screening criteria for admission into the study.

22. Eligibility

Participants are primarily Burundian children aged 9-15 years and their caregivers living in Nyanza Lac and Mabanda communes within the Makamba Provide, southern Burundi.
For schools to be eligible they had to be (1) public primary schools (école fondamental, EcoFo), (2) within the mentioned communes, (3) accessible by vehicle with no more than 30 min walk from where a vehicle could reach, (4) at least 1.5km a part from each other, (5) a minimum of 49 children in year 3, (6) school administration being willing to participate and accept the random allocation and potential classroom selection by coinflip as well as willing to welcome visitors, including research assistants and research coordinator for data collection, report information on physical education/activities common as well as enable TeamUp sessions and monitoring visits/observations to take place (for TeamUp schools only). TeamUp implementation schools had to have sufficient and adequate/safe space to conduct TeamUp sessions outside.
The school exclusion criteria (1) current psychosocial or mental health intervention implemented in the school, (2) school is less than 1.5 km away from other school selected for the research, and (3) school has less than 49 students enrolled in year 3. In case a selected school had one year 4 classroom, all children in the classroom were eligible to participate. Children in year 4 who are aged 9-15 years were eligible to participate in the study. If the school had more than one year 4 classroom, then group A was chosen. If <40 children per class, then group A and B were selected to reach the sample size. All children within the selected classes (i.e. entire grade 4, or group 4 A, or in a few cases also group 4 A and B) were allowed to join TeamUp sessions (unless in the control condition). Individual selection of children to participate was not possible due to ethical and logistical reasons, as well as due to the nature of TeamUp (group-based).

23. Recruitment
Children were invited to participate based on them being in the selected schools and classrooms. If no caregiver was able to come during baseline data collection days to provide consent or the child did not provide assent, the children and caregiver were not included. Figure 2 shows the CONSORT 2010 flow chart that will include the number of children being approached, number of eligible children identified, number of children agreeing to enter the trial, number of children refusing, and then by treatment arm: the number continuing through the trial, the number withdrawing, the number lost to follow-up and the numbers excluded / analysed, once the analyses are finalized and the statistician is unmasked to group allocation. For caregivers, teachers and facilitators, the sample size and loss-to-follow-up (flow chart information) will be reported in text.

Figure 2: CONSORT 2010 flow chart

Enrollment

Assessed for eligibility 72 schools

Excluded (n= 30) / Included (n=42)
- Not meeting inclusion criteria
- Declined to participate
- Other reasons

Randomly selected 36 out of 42, with 6 serving as potential reserves.

Randomized (n= 36)

Allocation

Allocated to intervention 12 TeamUp sessions (n=12 schools)
- Received allocated intervention (n=… children)
- Did not receive allocated intervention (give reasons) (n=… children)

Allocated to intervention 24 TeamUp sessions (n= 12 schools)
- Received allocated intervention (n=… children)
- Did not receive allocated intervention (give reasons) (n=… children)

Allocated to control group (n= 12 schools)
- Interviewed (n=… children)

Baseline

Allocated to intervention 12 TeamUp sessions (n=12 schools)
- Interviewed (n=… children)
- Received intervention (n= children)
- Did not receive allocated intervention (give reasons) (n=… children)

Allocated to intervention 24 TeamUp sessions (n= 12 schools)
- Interviewed (n=   children)
- Received intervention (n= children)
- Did not receive allocated intervention (give reasons) (n=… children)

Allocated to control group (n= 12 schools)
- Interviewed (n=… children)

Midline/Endline

- Interviewed (n=… children)
- Lost to follow-up (give reasons) (n=… children)

- Interviewed (n=… children)
- Lost to follow-up (give reasons) (n=… children)

- Interviewed (n=... children)
- Lost to follow-up (give reasons) (n=... children)

Follow-Up

- Interviewed (n=…children)
- Lost to follow-up (give reasons) (n=... children)

- Interviewed (n=… children)
- Lost to follow-up (give reasons) (n=... children)

- Interviewed (n=… children)
- Lost to follow-up (give reasons) (n=... children)

Analysis

- Analysed (n=... children)
- Excluded from analysis (give reasons) (n=... children)

- Analysed (n=... children)
- Excluded from analysis (give reasons) (n=... children)

- Analysed (n=... children)
- Excluded from analysis (give reasons) (n=... children)

24. Withdrawal / follow-up

Participants were free to withdraw from the study at any time without consequence. Details regarding loss to follow up will be provided in the CONSORT flow chart and reasons for loss of participants will be specified in the final report.

25. Baseline participants characteristics

The baseline characteristics of the participants will concern the total sample size (children and caregivers), and split out per arm. For children we will show the demographics as in table 1, and descriptives of age (mean and SD). We will not test for between arm differences on socio-demographic variables or baseline scores (Altman & Dore, 1991).

For the primary caregiver, we will show the demographics of mean age and SDs, as well as the distribution of male and female, relationship to child, highest education, occupation, head of household, frequency of receiving income, housing situation as well as if the person had ever been displaced, how many times and if so for what reason and for what duration. See format of table 2.

For the teachers and facilitators, we will show their basic demographics including age, gender and the years or experience. Facilitator data was only collected for the intervention arms. The demographics of teachers and facilitators will be reported in text. For teachers we will report on the years of experience of teaching, for facilitators on the months of experience in movement/sports based activities and months of previous experience with children.

Table 1
Children demographics at baseline in frequencies and percentages for total sample and split across trial arms

		Children 12 TU sessions (n=XXXX)	Children 24 TU sessions (n=XXXX)	Children control group (n=XXXX)	total (n=XXXX)
Gender	Male				
	Female				
Family composition	Number of children in household				
Country of birth	Burundi				
	Tanzania				
	Rwanda				
	DRC				
	other				
Separation from parents	Yes				
	No				
Reason of separation	Divorce				
	War (refugee)				
	Partial				

Repatriation (returning to Burundi)	No/never Yes, recent in 2023 Yes, in past 3 years Yes, within past 3-5 years Yes, more than 5 years ago				
Disability	None Physical disability Mental disabilities Physical and mental disabilities other				

Table 2
Caregiver demographics at baseline in frequencies and percentages for total sample and split across trial arms

		Caregivers 12 TU sessions (n=XXXX)	Caregivers 24 TU sessions (n=XXXX)	Caregivers control group (n=XXXX)	total (n=)
Gender	Male Female				
Relationship to child	Mother Father Brother/Sister (older) Uncle/Aunt (older) Grandfather/Grandmother Other family member Host family (non-family member) Other				
Education of caregiver	No schooling Informal Some of primary school Completed primary school Some of secondary school Completed secondary school University other No answer				
Occupation	Agriculture/Breeding Trade Tailor				

	Mechanic
	Vendor of products
	Fisherman
	Construction worker
	Transport sector
	Official job in public sector
	Private Sector Employee
	Retired
	Currently without income
	Housekeeper/Housemaid
	Other
	No answer
Head of household	Sole female (working age)
	Sole male adult (working age)
	male with female (working age)
	aged woman
	aged man
	man with women (both aged) together
	other
Frequency of receiving income	per day
	per week
	per month
	depending on the season
	other
	No answer
Housing situation	renting
	owner
	living with someone for free
	other
	No answer
Reason of displacement	conflict
	war
	natural disasters
	Poverty
	Land conflict
	Others
	No answer

Analysis

26. Outcome definitions

All outcomes were measured at baseline, endline and 6-months follow-up. At midline (for 24 session arm only), we also assessed the psychosocial wellbeing (SCWBS) and the self-identified problems (PSYCHLOPS-inspired). See table 3 for details.

The primary outcome measure is psychosocial wellbeing (Stirling Children's Wellbeing Scale, SCWBS, 12 items, plus 3 additional social desirability items).

The secondary outcome measures are:
- Children's Revised Impact of Event Scale (CRIES-8) measuring traumatic stress (8 items)
- Multidimensional Students Life Satisfaction Scale (MSLSS) with subscale School (9 items), assessing satisfaction and attitude towards school, and subscale Friends (8 items), assessing friendships
- Child-report KIDSCREEN-10 measuring health-related quality of life assessing subjective wellbeing (10 items)
- Child-report KIDSCREEN-52 subscale physical activities and health (5 items) measuring physical health
- PSYCHLOPS-inspired questions, i.e. impact of self-defined problems (2 open questions to define problems, and ratings of these, thus 2 items)
- Children's Hope Scale measuring hope (6 items)
    o *also a hypothesized mediator*
- Emotional Regulation Questionnaire (ERQ) assessing emotional regulation (10 items)
    o *also a hypothesized mediator*
- Short Moods and Feelings Questionnaire (SMFQ-13) measuring depression symptoms (13 items)
- TeamUp-specific Mechanisms of Action scale (self-developed, 30 items), assessing children's sense of safety (5 items), social connectedness (6 items), sense of agency (6 items), emotional regulation (9 items) and interoception (4 items)
    o *also a hypothesized mediator*

Additionally, we collected the following implementation quality outcomes:
- Attendance, measured for each of the sessions as present / absent
- Fidelity of facilitators (27 items, including 12 individual-level fidelity and 15 team-level fidelity times, in addition to 6 TeamUp-specific competencies), measured as not done/partly done/well done per session and as a percentage over all sessions
- WeACT (9 items of the original 13-item tool, Jordans et al., 2021; Jordans et al., 2022), assessed at recruiting stage and during practice session weeks

Other outcomes:
- Caregiver-reported child wellbeing (parent-report KIDSCREEN-10, 10 items)
- Facilitator wellbeing (Warwick-Edinburgh Mental Wellbeing Scales, WEMWBS, 14 items), resilience (Connor-Davidson Resilience Scale, CD-RS 10 items) and satisfaction and burden (self-developed tool, see Bleile et al., 2021), all assessed pre- and post-implementation (baseline and endline)
- Teacher-reported classroom dynamics (inspired by the Student-Teacher Relationship Scale, yet self-developed; 5 items), reporting on their classroom overall as a group, assessed at baseline and endline
- Teacher-reported children's behavior and performance in class (adapted and shorted Classroom Performance Survey; 9 items), reporting on 5 index children in their classroom, assessed at baseline and endline

Note: Developers and references of outcome measures are mentioned and cited in the protocol.

Table 3
Study measures and timepoints

Measures	Baseline	Midline*	Endline	Follow-Up
Children				
SCWBS	X	X	X	X
CRIES-8	X		X	X
MSLSS	X		X	X
child-report KIDSCREEN-10	X		X	X
child-report KIDSCREEN-52 subscale physical activities and health	X		X	X
PSYCHLOPS-inspired questions	X	X	X	X
Children's Hope Scale	X		X	X
ERQ	X		X	X
SMFQ-13	X		X	X
TeamUp-specific Mechanisms of Action scale	X		X	X
Caregivers				
Parent-report KIDSCREEN-10	X		X	X
Teachers				
classroom dynamics	X		X	
children's behavior and performance in class	X		X	
Facilitators				
WEMWBS	X		X	
CD-RS	X		X	
Satisfaction and burden	X		X	

* Midline was only collected for the 24-TeamUp-session arm and the control arm, as this was collected at the same time as endline of the 12-TeamUp session arm, see Figure 1.

27. Analysis methods

Prior to the main analyses, participants' attendance will be described by the median, 25th and 75th percentiles and range of the number of sessions attended, as well as the number and proportion of participants who were considered to have adhered to the treatment, for each arm. This equates to attendance of equal to or over 75% of the intended TeamUp session for the 12 and 24 session arms. As this will unmask the statistician, computing of the adherence variable will be done by another independent statistician/lead researcher. This variable will be needed early in the analysis in order to assess whether post-randomisation variables, such as this adherence variable, are related to missing follow-up data.

The numbers and proportions of participants with missing data for each baseline, primary and secondary variable will be summarised overall, and by arm and time point. The baseline characteristics of those missing follow up (midline, endline, and follow-up) will be compared to those with complete follow-up using descriptive statistics and if possible, depending on how many cases, a logistic predictor of missingness model (Fairclough, 2010). The number and proportion actively withdrawing from the trial and reasons for withdrawal will be summarised overall and by treatment group separately from those that are passively lost to follow-up.

We will follow the steps below for our analyses.

Step 1: Assumption check and descriptive statistics
We will conduct assumptions checks and provide descriptive statistics:
- Distribution checks of continuous measures by means of QQ plots and histograms

14

- o Mean and standard deviations for all symmetric (non-skewed) distributed measures
- o Median, 25th and 75th quartiles for skewed distributed measures
- o Frequencies and proportions (percentage) of categorical measures
- o Table with descriptive statistics for the total sample, per intervention group and per time point, in the format of table 4.
  - ▪ This Information split by school will be presented in a supplementary Table.

Table 4
Means, standard deviations, mean difference scores, and 95% CIs.

		Baseline	Midline	Endline	Follow-up	M_{diff} BL-EL	95% CI BL-EL	M_{diff} BL-FU	95% CI BL-FU
Psychosocial wellbeing	TU12								
	TU24								
	control								
	Total								
Traumatic stress	TU12		n/a						
	TU24								
	control								
	Total								
Attitude towards school	TU12		n/a						
	TU24								
	control								
	Total								
Friendships	TU12		n/a						
	TU24								
	control								
	Total								
Health-related quality of life	TU12		n/a						
	TU24								
	control								
	Total								
Physical health	TU12		n/a						
	TU24								
	control								
	Total								
Hope	TU12		n/a						
	TU24								
	control								
	Total								
Emotional regulation	TU12		n/a						
	TU24								
	control								
	Total								
Depression symptoms	TU12		n/a						
	TU24								
	control								
	Total								
PSYCHLOPS	TU12								
	TU24								
	control								
	Total								

Outcome	Timepoint				
Mechanisms of Action (MoA)	TU12	n/a			
	TU24				
	control				
	Total				
Sense of Safety (MoA)*	TU12	n/a			
	TU24				
	control				
	Total				
Social connectedness (MoA)*	TU12	n/a			
	TU24				
	control				
	Total				
Sense of agency (MoA)*	TU12	n/a			
	TU24				
	control				
	Total				
Emotional regulation (MoA)*	TU12	n/a			
	TU24				
	control				
	Total				
Interoception (MoA)*	TU12	n/a			
	TU24				
	control				
	Total				
Caregiver-reported child health-related quality of life	TU12				
	TU24				
	control				
	Total				
Teacher-reported classroom dynamics	TU12			n/a	
	TU24				
	control				
	Total				
Teacher-reported child behaviour/performance in class	TU12			n/a	
	TU24				
	control				
	Total				
Facilitator wellbeing	TU12				
	TU24				
	control		n/a	n/a	n/a
	Total				
Facilitator resilience	TU12				
	TU24				
	control		n/a	n/a	n/a
	Total				
Facilitator satisfaction and burden	TU12				
	TU24				
	control		n/a	n/a	n/a
	Total				

* Note: these are subscales of the Mechanisms of Action (MoA) scale

Step 2: Main analysis of primary and secondary outcomes

For the **primary analysis**, the outcome is the total score of psychosocial wellbeing (Stirling Children's Wellbeing Scale, SCWBS) at endline and 6 months post randomisation adjusted for SCWBS at baseline and account for hierarchical clustering of children at the level of school. Specifically, we will fit a three-level random intercept linear regression model considering observations from baseline, endline and 6 months at level 1, children at level 2 and schools at level 3. The model will include the total score of SCWBS at endline and 6 months as the outcome variables with the three-arm treatment group (control; dosage A; dosage B, baseline rate of SCWBS as explanatory variables. A time by treatment interaction will be included to allow the effect to differ at each time point.

Secondary outcomes (traumatic stress, satisfaction and attitude towards school, friendships, health-related quality of life, physical health, impact of self-defined problems, hope, emotional regulation, depression symptoms, as well as (self-developed tools) Mechanisms of Action, including the subscales of sense of safety, social connectedness, sense of agency, emotional regulation and interoception) will be assessed with a similar methodology for the primary outcomes, using generalized linear mixed models with the appropriate link function/distribution. Normally distributed outcomes will use an identity link function, binary variables would use a logit link function, and Poisson (count) variables would use a log link function. We will use all available time points in the model and extract differences at endline and 6 months post-intervention.

Specifically:
- o If the interaction effect is significant, post-hoc comparisons will be examined for differences between 12 and 24 sessions of TU at endline and follow-up
- o The primary comparison point is endline (immediately after the end of the intervention), secondary comparison point is 6-month follow up
- o We test for superiority of 24 TeamUp sessions over 12 TeamUp sessions, for the primary outcome measure as well as secondary measures at endline
- o If no superiority is shown, a non-inferiority test will be employed, testing if 12 TeamUp sessions do not perform worse than 24 TeamUp sessions at endline, on the primary outcome as well as secondary outcomes
- o No statistical significance tests or confidence intervals will be calculated for the difference between randomised groups on any participant level baseline variables. The randomisation of participants to intervention groups means that any imbalance over all measured and unmeasured baseline characteristics is by definition due to chance (Altman & Dore, 1991).

Step 3: Secondary analysis comparing TU total to control

We will test with a similar approach as described above whether TeamUp leads to increased wellbeing compared to control, *regardless of the number of sessions*. Thereby, we will merge 12 and 24 TeamUp session arms together, after checking the assumption of homogeneity of variances with a scatterplot and Levene's / Bartlett's test.

Step 4: Secondary analysis comparing 12 TU sessions to control

We will test with a similar approach as described above the difference between baseline and midline/endline (February assessment), across all three arms on the primary outcome measure, expecting that the psychosocial wellbeing of children at endline of 12 TeamUp sessions is equivalent to the midline point of 24 TeamUp sessions (thus after 12 TeamUp sessions), and both to outperform the control group – given both arms of children participating in TeamUp had been provided 12 TeamUp sessions by that time point.

Step 5: Exploratory mediation models

Single mediator models: We will test for the effect of the TeamUp (combining 12 and 24 TeamUp session arms) on each one of the multiple continuous mediators (MoA, hope, emotional regulation) with the use of linear regression models. We will choose a priori a p-value of lower than 0.15 to select the appropriate mediators to include in our final mediation analysis. Multiple mediators require independence with each other. We will test independence by examining partial correlations between our mediators after accounting for treatment allocation.

Full Mediation Model: For the endline and 6 months showing a significant intervention effect on children's psychosocial wellbeing, a sequential mediation model will be used to estimate the total effect, indirect effects (IE), and direct effects (DE) of TeamUp on child psychosocial wellbeing (SCWBS) at 6-months follow-up, see Figure 3, model 1. To quantify the magnitude of mediation, the proportion of the effect mediated by putative mediators (NIE/[NDE + NIE]) will be given. All analyses will be estimated using bootstrapping (500 replications) to recover the correct SEs for direct and indirect effects.

Specifically:
        Predictor= arm (control vs, TU 12+24)
        Mediator 1= MoA at endline
        Mediator 2= hope at endline
        Mediator 3= emotional regulation at endline
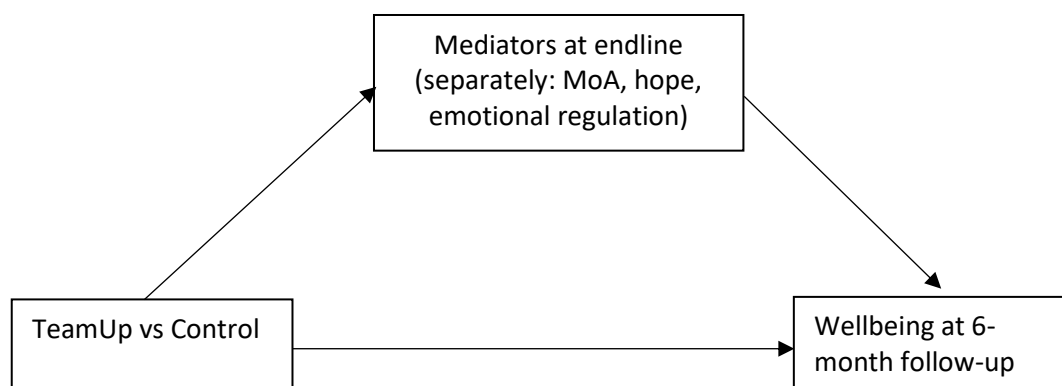   Outcome= wellbeing at follow up

A separate mediation analysis will then test attendance of the combined 12 and 24 session groups as a predictor of outcomes (psychosocial wellbeing at 6-months follow-up), and MoA as a mediator, see Figure 3, model 2.
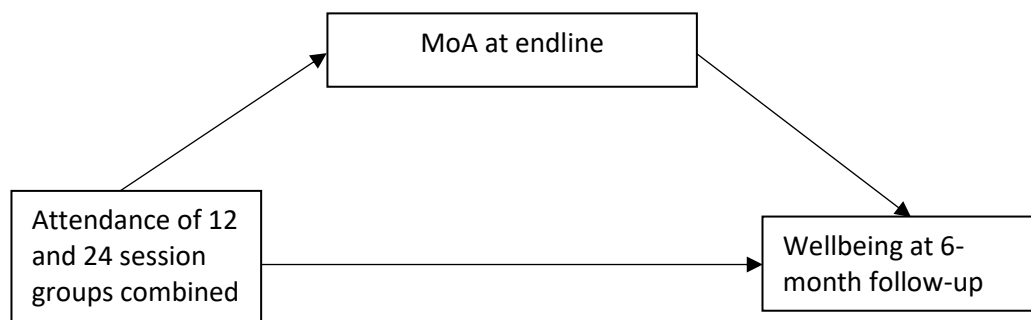
Specifically:
        Predictor= attendance (in percentage)
        Mediator 1=MoA
        Outcome= wellbeing

Figure 3. Directed Acyclic Graphs

Model 1

Model 2

```
                    ┌─────────────────────┐
                    │   MoA at endline    │
                    └─────────────────────┘
                  ↗                          ↘
┌──────────────────┐                    ┌──────────────────┐
│ Attendance of 12 │                    │  Wellbeing at 6- │
│ and 24 session   │ ─────────────────→ │ month follow-up  │
│ groups combined  │                    │                  │
└──────────────────┘                    └──────────────────┘
```

28. Missing data

Although baseline data should be complete, there may be some limited missing data. Descriptive baseline summaries will then be presented as complete case. The proportion of missing data will be summarised per outcome. If any of the baseline measures are found to relate to missing primary outcome at midline, endline or 6 months, we will consider adjusting for them in models for the primary outcome. To allow for this, any baseline measure considered as a covariate in the main model would best be imputed to a full single dataset. Missing baseline covariate data will therefore be imputed (MI, 5 datasets).

For outcome measures where there are published methods for dealing with missing items, these will be applied. Otherwise, we will prorate missing items only when there are no more than 20% missing items (i.e. for a ten item questionnaire, prorate only where one or two items are missing) by replacing the missing item values with the median value of the complete items for each individual. If after prorating there are still missing total questionnaire scores at baseline, these will be imputed as described above.

29. Additional analyses

We will perform a subgroup analysis of participants in the trial by gender, repatriated, separation from caregiver, only including children with high scores (≥12) on the SMFQ, indicating depression symptoms.

We will conduct the Per Protocol analysis, as defined above (see page 7).

We will conduct sensitivity analyses excluding children with high social desirability scores on the SCWBS, as per guidance of SCWBS.

We will run subgroup analyses on the primary outcome:
- o excluding children with high social desirability scores (as per guidance of SCWBS)
- o split by child gender
- o split by child repatriated status yes/no
- o split by children having been separated from their caregivers yes/no
- o only including children with high scores (≥12) on the SMFQ, indicating depression symptoms
- o We will conduct sensitivity analyses depending on anomalies found during data analyses (post-hoc).

We will conduct additional analyses to test;

o caregiver-reported child wellbeing, using the same model as the primary analysis with interaction effects of group x time, including the three arms and baseline, endline and follow-up.

  o First, we will include all caregiver-report child wellbeing (parent-report Kidscreen-10), even for cases where caregivers across time points differed.

  o Secondary, we will assess the caregiver-report child wellbeing for all children for which we have the same, matching BL-EL-FU caregiver.

o teacher-reported outcomes of classroom dynamics as well as index children's performance will be assessed, using the same model as the primary analysis but with baseline as the covariate.

o facilitator wellbeing, comparing baseline to endline, comparing facilitators implementing 12 TeamUp sessions to facilitators implementing 24 TeamUp sessions. The control group did not have any facilitator data.

30. Harms

Adverse Events (AEs) and Severe Adverse Events (SAEs) will be tabulated by event type and body system classification by time and randomised group. Events will also be tabulated as number of events and number of people having an event. Each table will detail the number of children that were still in the trial at the time points by randomisation group. We will consider calculating adverse and serious adverse event incidence rates and 95% confidence intervals by group. If any adverse events are selected as being of particular interest they will be further summarised; outlining the severity (mild, moderate, severe) and if classified as serious the expectedness (expected, unexpected) of the event. Frequencies of adverse events will be reported as occurred. We will follow the adverse events procedure outlined in our Data Management Plan.

31. Statistical software

The software used for the analyses will be JASP (version 16.2) and R Studio (version 4.4.3, mediation package, Tingley et al., 2014).

32. References

Altman, D. G., & Dore, C. J. (1991). Baseline comparisons in randomized clinical trials. *Statistics in Medicine*, *10*(5), 797-799.

Bareka, T., Panhofer, H., & Rodriguez Cigaran, S. (2019). Refugee children and body politics. The embodied political self and dance movement therapy. *Body, Movement and Dance in Psychotherapy*, *14*(2), 80-94.

Bleile, A. C. E, Koppenol-Gonzalez, G. V., Verreault, K., Abeling, K., Hofman, E., Vriend, W., ... & Jordans, M. J. (2021). Process evaluation of TeamUp: a movement-based psychosocial intervention forrefugee children in the Netherlands. *International journal of mental health systems*, *15*(1), 1-16.

Bleile, A. C., Koppenol-Gonzalez, G. V., Orech, B., Verreault, K., & Jordans, M. J. (2024). Evaluating a movement-based mental health promotion intervention for refugee children in Uganda: A quasi-experimental study. *Journal of Clinical Child & Adolescent Psychology*, 1-16.

Dieterich-Hartwell, R. (2017). Dance/movement therapy in the treatment of post traumatic stress: A reference model. *The arts in psychotherapy*, *54*, 38-46.

Dieterich-Hartwell, R., & Koch, S. C. (2017). Creative arts therapies as temporary home for refugees: Insights from literature and practice. *Behavioral Sciences*, *7*(4), 69.

Fairclough, D. L. (2010). *Design and analysis of quality of life studies in clinical trials*. Chapman and Hall/CRC.

Gamble, C., Krishan, A., Stocken, D., Lewis, S., Juszczak, E., Doré, C., ... & Loder, E. (2017). Guidelines for the content of statistical analysis plans in clinical trials. *Jama*, *318*(23), 2337-2343. doi:10.1001/jama.2017.18556

Hobfoll, S. E., Watson, P., Bell, C. C., Bryant, R. A., Brymer, M. J., Friedman, M. J., ... & Maguen, S. (2007). Five essential elements of immediate and mid–term mass trauma intervention: Empirical evidence. *Psychiatry: Interpersonal and Biological Processes, 70*(4), 283-315.

Inter-Agency Standing Committee (IASC) (2007). IASC Guidelines on Mental Health and PsychosocialSupport in Emergency Settings. Geneva: IASC.

Jordans, M. J. D., Coetzee, A., Steen, H. F., Koppenol-Gonzalez, G. V., Galayini, H., Diab, S. Y., ... & Kohrt, B. A. (2021). Assessment of service provider competency for child and adolescent psychological treatments and psychosocial services in global mental health: evaluation of feasibility and reliability of the WeACT tool in Gaza, Palestine. *Global Mental Health*, *8*, e7.

Jordans, M. J. D., Steen, F., Koppenol-Gonzalez, G. V., El Masri, R., Coetzee, A. R., Chamate, S., ... & Kohrt, B. A. (2022). Evaluation of competency-driven training for facilitators delivering a psychological intervention for children in Lebanon: a proof-of-concept study. *Epidemiology and psychiatric sciences*, *31*, e48.

Jordans, M. J., Koppenol-Gonzalez, G. V., Bleile, A. C., Orech, B., Brian, A., & Verreault, K. (2024). Follow-up and mediation outcomes of a movement-based mental health promotion intervention for refugee children in Uganda. *Journal of Clinical Child & Adolescent Psychology*, 1-12.

Molero-Calafell, J., Burón, A., Castells, X., & Porta, M. (2024). Intention to treat and per protocol analyses: differences and similarities. *Journal of Clinical Epidemiology, 173,* 111457.

Pierce, L. (2014). The integrative power of dance/movement therapy: Implications for the treatment of dissociation and developmental trauma. *The Arts in Psychotherapy*, *41*(1), 7-15.

Sullivan, T. R., White, I. R., Salter, A. B., Ryan, P., & Lee, K. J. (2018). Should multiple imputation be the method of choice for handling missing data in randomized trials?. *Statistical methods in medical research*, *27*(9), 2610-2626.

TeamUp. (2020a). *HandBook For facilitators of structured movement-based activities in safe spaces, schools and other community environments.* War Child, Save the Children and UNICEF Netherlands, Amsterdam.

TeamUp. (2020b). *GameBook For facilitators of structured movement-based activities in reception centres an schools.* War Child, Save the Children and UNICEF Netherlands, Amsterdam.

The World Bank (2023). Where we work: Burundi Overview. Available at: https://www.worldbank.org/en/country/burundi Accessed on 3 July 2023.

Tingley, D., Yamamoto, T., Hirose, K., Keele, L., Imai, K. (2014). Mediation: R package for causal mediation analysis. *Journal of Statistical Software, 59* (3). Doi: 10.18637/jss.v059.i05

Tortora, S. (2005). *The dancing dialogue: Using the communicative power of movement with young children*. Redleaf Press. 10 Yorkton Court, St. Paul, MN 55117-1065.

UNHCR (2019). Briefing Notes: UNHCR and partners seek US$296 million for Burundi refugee crisis. Available at: https://www.unhcr.org/news/briefing-notes/unhcr-and-partners-seek-us296-million-burundi-refugee-crisis Accessed on 3rd July 2023.

Ventevogel, P. (2015). The effects of war: local views and priorities concerning psychosocial and mental health problems as a result of collective violence in Burundi. *Borderlands of mental health: Explorations in medical anthropology, psychiatric epidemiology and health systems research in Afghanistan and Burundi*, 49.

Ventevogel, P., Ndayisaba, H., & Van de Put, W. (2011). Psychosocial assistance and decentralizedmental health care in post-conflict Burundi. *Borderlands of mental health*, 211.

War Child Holland (2019). *Data Management Plan, Research &Development Department*. War Child Holland, Amsterdam.

White, I. R., & Thompson, S. G. (2005). Adjusting for partially missing baseline measurements in randomized trials. *Statistics in medicine*, *24*(7), 993-1007.

Timeline

Date	Action	Responsible
28 October 2024	SAP first draft to GK and MJ for input	AB
16 November	Feedback provided on first draft	MJ; GK
25 November	Incorporating feedback to create second draft	AB
26 November	Second SAP draft discussed	MJ, GK, AB
9 December	Third SAP draft shared with GK and MJ	AB
10-17 December	Feedback received via email	MJ, GK, AB
18 December	Fourth SAP draft shared with GK and MJ	AB
20-29 December	Revision of SAP	GK; MJ
7 January 2025	Final SAP shared with all co-authors	MJ and AB
8-15 January	Revisions made based on input from all co-authors	All co-authors
8 January	Data cleaning process started	AB
20-23 January	Create final SAP version and have it signed by all co-authors	AB; GK; MJ
21 January	Completed data cleaning	AB
23 January	Raw and cleaned data to GK	AB
	Datasets prepared for analysis	GK
	Primary analyses finalized	GK
	Review analysis	IB
10 March	All analyses finalized	GK
31 Oct	Primary publication	MJ